



How to Assess Your CURE: A Practical Guide for Instructors of Course-Based Undergraduate Research Experiences[†]

Erin E. Shortlidge* and Sara E. Brownell
School of Life Sciences, Arizona State University, Tempe, AZ 85201

Integrating research experiences into undergraduate life sciences curricula in the form of course-based undergraduate research experiences (CUREs) can meet national calls for education reform by giving students the chance to “do science.” In this article, we provide a step-by-step practical guide to help instructors assess their CUREs using best practices in assessment. We recommend that instructors first identify their anticipated CURE learning outcomes, then work to identify an assessment instrument that aligns to those learning outcomes and critically evaluate the results from their course assessment. To aid instructors in becoming aware of what instruments have been developed, we have also synthesized a table of “off-the-shelf” assessment instruments that instructors could use to assess their own CUREs. However, we acknowledge that each CURE is unique and instructors may expect specific learning outcomes that cannot be assessed using existing assessment instruments, so we recommend that instructors consider developing their own assessments that are tightly aligned to the context of their CURE.

INTRODUCTION

There have been national recommendations to integrate research experiences into the undergraduate biology curriculum (1, 49). While independent research in a faculty member’s lab is one way to meet this recommendation, course-based undergraduate research experiences (CUREs) are a way to scale the experience of doing research to a much broader population of students, thereby increasing the accessibility of these experiences (4). In contrast to traditional “cookbook” lab courses where students complete a pre-determined series of activities with a known answer, CUREs are lab courses where students work on novel scientific problems with unknown answers that are potentially publishable (3, 11). These research experiences have been shown to benefit both students and the instructors of the CUREs (45, 54, 55). As such, CUREs are growing in popularity as an alternative to the traditional laboratory course, particularly in biology lab classes (3, 10, 12, 36, 37, 51, 54, 62). The in-class research projects often last the duration of a semester or quarter, although sometimes they are implemented as shorter modules. CUREs vary in topics, techniques, and research questions—with the common thread being that the scientific questions addressed are

novel, and perhaps most significantly, of interest to the scientific research community beyond the scope of the course. CUREs meet national calls for reforming biology education and may provide unique learning outcomes for students, as emphasized in Vision and Change: “learning science means learning to do science” (1).

Studies have reported various student outcomes resulting from CUREs including: increased student self-confidence (6, 40), improved attitudes toward science (35, 36), ability to analyze and interpret data (10, 13), more sophisticated conceptions of what it means to think like a scientist (10), and increased content knowledge (34, 46, 62). Although there is general consensus that CUREs can have a positive impact on students, it is often unclear what specific aspect of a CURE leads to a measured outcome (17). Further, given the uniqueness of each individual CURE, instructors and evaluators of CUREs may struggle to identify how to effectively assess particular elements of their CURE. Further, a combination of assessment techniques may yield the most holistic understanding of course outcomes, especially for CUREs that are already being implemented (3, 29).

While an ideal assessment of outcomes from any experimental pedagogy would include an analysis of a matched comparison course, logistical reasons may likely prevent an instructor from executing a randomized controlled study or even a quasi-experimental design format allowing a comparison of students in a CURE with students in a non-CURE course (but see 13, 31, 36 for examples). Further, finding an appropriate comparison group of students can be difficult; in particular, if students choose to take the CURE, then there is a possible selection bias (13). However, one can

*Corresponding author. Mailing address: Biology Department, 1719 SW 10th Ave. SRTC 246, Portland State University, Portland, OR 97219. Phone: 503-725-9305. Fax: 503-725-3888. E-mail: eshortlidge@pdx.edu.

[†]Supplemental materials available at <http://asmscience.org/jmbe>

account for known differences and similarities in students by using multiple linear regression models that control for student demographics and ability such as incoming grade point average (GPA) (e.g., 59). Although these are all factors to take into consideration, in this article, our aim is to provide instructors with fundamental guidelines on how to get started assessing their own CUREs, without necessarily needing a comparison group.

In assessing their CURE, we propose that instructors first begin with their intended learning outcomes for the CURE. Once an instructor's learning outcomes have been identified, they can align assessments with those outcomes. We suggest that instructors browse existing assessments that may align with their learning outcomes to see whether any present appropriate ways to evaluate their CURE. If not, instructors should consider designing their own assessment situated in the specific context of their CURE, which may require collaboration with an education researcher or group of researchers with expertise in assessment. Finally, instructors need to critically evaluate the results of the assessment, being cautious in their interpretation. Taken together, these steps will provide a "best practices" model of how to effectively assess CURE learning environments (Fig. 1).

Step 1: Identify learning outcomes

The first step in evaluating a CURE is to identify learning goals so that an instructor can assess how successful the CURE was at attaining those learning goals. There are a number of resources available to help instructors identify and establish learning goals (33) and design a course based on these learning goals, i.e., 'backward design' (22, 64). We use the terms "learning goals" and "learning outcomes" as discussed by Handelsman et al. in *Scientific Teaching* (2004): "Learning goals are useful as broad constructs, but without defined learning outcomes, goals can seem unattainable and untestable. The outcome is to the goal what the prediction is to the hypothesis" (33). Thus, the first question to ask is not, "How can I assess my CURE?", but "What learning

outcomes do I want to measure?" Learning outcomes can vary, ranging from technical skills (e.g., "students will be able to micropipette small amounts of substances accurately"), to content knowledge (e.g., "students will be able to explain the steps of the polymerase chain reaction"), to high-level analytical skills (e.g., "students will be able to adequately design a scientific experiment"). More general learning goals for a CURE may also include affective gains such as self-efficacy or improved attitude toward science (6, 36). These affective gains may be more difficult for traditionally trained biologists to measure, or to evaluate the results of, particularly if one is less familiar with the theoretical frameworks that define these constructs. Collaboration with experts in education to understand affective gains may be particularly appropriate if these are the anticipated outcomes for one's CURE. These presented learning outcomes are intended to serve as examples of the diversity of conceivable student outcomes from a CURE but reflect only a fraction of those possible. Frameworks to identify learning outcomes from CUREs have been developed elsewhere (e.g., 11, 16) and could be used in conjunction with the present article as a starting point for CURE assessment.

While it is possible for CUREs to lead to gains in various domains, instructors may want to focus the learning goals on those with potentially measurable learning outcomes that are either not feasible in a lecture course, or are best-suited for a lab course. For example, any biology course can cover content, but in addition to other possible gains afforded by the CURE format, a lab course focusing on novel data is uniquely positioned to teach students about the *process* of science or perhaps the importance of repeating experiments. However, if the CURE is a required course in a department or if CUREs are being taught parallel to traditional lab courses, there may be already-established departmental learning goals and specific outcomes that must be targeted by the course.

Step 2: Select an assessment aligned with your learning outcomes

Once the anticipated learning outcomes for the CURE have been identified, the next step is to find an assessment strategy that aligns with the learning goals and anticipated learning outcomes. For some of the more common learning outcomes from CUREs, assessment instruments may either already exist or have been developed specifically for CUREs. These are sometimes referred to as "off-the-shelf" assessments because they have previously been published and instructors could in theory grab one "off the shelf" and administer it in the CURE classroom. However, it is important to consider how well these assessment instruments measure an instructor's specific intended learning outcomes. Tight alignment of the assessment instrument with the desired learning outcomes is essential to accurately interpret CURE results. We further encourage instructors to critically evaluate these instruments in terms of:

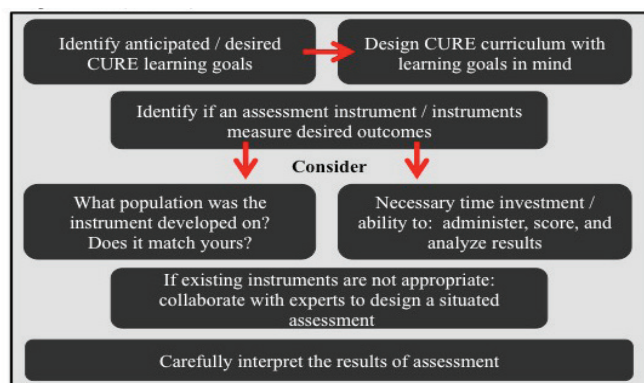


FIGURE 1. Guide to assessing course-based undergraduate research experiences (CUREs).

administration (e.g., How much class time does it require to administer? Is it a pre-post-course comparison?), time required to score the assessment (e.g., multiple choice, which can be auto-graded vs. open-ended responses, which need to be evaluated with a rubric), and what validation has been conducted on the instrument and how appropriate it is for an instructor's specific population of students (e.g., has the instrument been previously administered to high school students but not to undergraduates?).

Previously developed assessments. Table 1 outlines assessment instruments that instructors could potentially use to evaluate their CURE, ordered by their primary aims. The table includes details of the format of the assessment, ease of administration and grading, and the population(s) that the instrument has been developed on. While we are not intending for this list to be comprehensive, and there are likely other assessment instruments that can be used to assess CUREs, we hope that this could be helpful for CURE instructors who are at the beginning stages of thinking about assessment.

Included in the table are primary references for each instrument so instructors can find more information on the process of the development of the instrument as well as the efforts made by the assessment developers to ensure that the instruments produce data that has been shown to be valid and reliable (25, 47, 58). There are a set of best practices standards for educational and psychological measures outlined by the American Educational Research Association (2), and assessment instruments ideally adhere to these standards and provide evidence in support of the validity and reliability of the resulting data. It is important to note that no assessment instrument is generally validated—it is only valid for the specific population on which it was tested. An assessment instrument that was developed on a high school population may not perform the same way on a college-level population, and even an assessment instrument developed on a population of students at a research-intensive institution may not perform the same way on a community college student population. There is not a “one size fits all” for assessment, nor is there ever a perfect assessment instrument (5). There are pros and cons to each depending on one's specific intentions for using the assessment instrument, which is why it is critical for instructors to judiciously evaluate the differences among instruments before choosing to use one.

If no existing assessment fits, then design your own. Existing assessment instruments may not be specific enough to align with an instructor's anticipated learning outcomes. Instructors may want to measure learning outcomes that are specific to the CURE, and using an instrument that is not related to the specific context of the CURE may not be able to achieve that. We recommend that instructors consider working with education researchers to design their own assessments

that are situated in the context of the CURE (e.g., 38), and/or use standard quizzes and exams as a measure of expected student CURE outcomes (e.g., 10). The choices instructors make will depend on the intention of their assessment efforts: is the intent to make a formative or summative assessment? What does the instructor intend to learn from and do with the measured outcome data? For example, do they wish to use the results to advance their own knowledge of the course success, for a research study, or a programmatic evaluation?

Step 3: Interpret the results of the assessment

Once instructors administer an assessment of their CURE, it is important to be careful in interpreting the results. In a CURE, students are often doing many different things and it is difficult to attribute a learning gain to one particular aspect of the course (16). Further, a survey that asks students about how well they think they can analyze data is measuring *student perception* of their ability to analyze data (e.g., 12), which could be different than their actual ability (e.g., 38) or the instructor's perception of that student's ability to analyze data (e.g., 55). Thus, it is important that instructors not try to overgeneralize the results of their assessment, and that they are aware of the limitations of student self-reported gains (9, 40). Yet, student perceptions are not always limitations, as student self-report can be the best way to measure learning goals such as confidence, sense of belonging, and interest in pursuing research—here it is appropriate to document how a student feels (15). Further, the instructor may want to know what the student *thinks* they are gaining from the course. For example, if an instructor's expected learning outcome is for students to learn to interpret scientific figures, they could work to answer the question using a multi-pronged approach, measuring student perception of ability paired with a measure of actual ability. To achieve this, an instructor could use or design an assessment that asks students to self-report on their perceived ability to interpret scientific graphs. The instructor could then pair the self-report instrument with an assessment testing their actual ability to interpret scientific graphs. Using this approach, an instructor could learn whether there is alignment between what the instructor thinks the students are learning, what the students think they are learning, and whether the students are actually learning the skill. Thus, the attributes and limitations of assessment instruments and strategies are dependent on both the learning outcomes one wants to measure and the conclusions one wants to draw from the data.

Putting the steps in action: An example of alignment of goals and assessment

Here we present guiding questions for instructors to ask when approaching an assessment instrument. These

TABLE 1.
Assessment instrument table.

Assessment Instrument Name	Acronym	Primary Aim	Secondary Aim	Validation Population	Answer Type	How Administered	Number of Items / Time to Administer	Scoring	Citation
Blooming Biology Tool	BBT	Cognitive skills	Critical thinking	University of Washington	Open-ended	In class or online	NA	Moderate – rubric	(20)
California Critical Thinking Skills Test	CCTST ^a	Cognitive skills	Critical thinking	California State University Fullerton	Multiple choice	Online	45 min	Easy	(28)
Critical Thinking Assessment Test	CAT ^a	Cognitive skills	Critical thinking	Multiple populations	Open-ended	In class	15 open-ended items	Moderate – scoring guide	(59)
Study Process Questionnaire	R-SPQ-2F	Cognitive skills	Deep and surface learning	UG in Hong Kong	Likert-type	In class	20 items	Easy	(6)
Networking Scale ^b		Communication	Networking	University of Pittsburgh	Likert-type	Online	5 items	Easy	(34)
Perceived Cohesion Scale	PCS	Community and collaboration	Sense of belonging	Multiple ages & populations	Likert-type	In class	6 items	Easy	(7)
Torrance Tests of Creative Thinking	TTCT/ATTA	Creativity	Creativity	Multiple ages & populations	Open-ended		15 min	Moderate – scoring guide	(63)
Environmental Attitudes Inventory	EAI	Environmental Awareness	Environmental Attitudes	Multiple ages & populations	Likert-type	Online	24 or 72 items	Easy	(50)
New Ecological Paradigm Scale	NEP	Environmental awareness	Environmental awareness	Multiple ages & populations	Likert-type		15 items	Easy	(26)
Grit Scale	Grit	Grit	Perseverance, self-control, passion	Multiple ages & populations	Likert-type	In class	12 items	Easy	(25)
Views About Sciences Survey	VASS	Nature of science	Nature of science	Multiple ages & populations	Contrasting alternatives design (multiple choice)	In class	30 items	Easy	(32)
Views on the Nature of Science	VNOS-C	Nature of science	Nature of science	Multiple ages & populations	Open-ended	In class	45–60 min	Difficult – inter-rater & interviews	(44)
Project Ownership Survey	POS	Ownership	Project ownership	Multiple UG populations	Likert-type	Online	16 items	Easy	(33)
Career Decision Making Survey – Self Authorship	CDMS-SA	Ownership	Self-authorship	Multiple populations	Likert-type		18 items	Easy	(19)
Laboratory Course Assessment Survey	LCAS	Perceptions of biology	Collaboration, discovery and relevance, iteration	Multiple UG populations	Likert-type	Online	17 items	Easy	(18)

TABLE 1.
Continued.

Assessment Instrument Name	Acronym	Primary Aim	Secondary Aim	Validation Population	Answer Type	How Administered	Number of Items / Time to Administer	Scoring	Citation
Colorado Learning Attitudes about Science Survey	CLASS-Bio	Personal gains in context of science	Attitudes about discipline specific science/problem solving	University of British Columbia, University of Colorado Boulder	Likert-type	Online pre/post	31 items	Easy	(54)
Classroom Undergraduate Research Experience	CURE	Personal gains in context of science	Attitudes about science		Likert-type	Online pre/post	Adaptable, 15 min	Easy – done by author	(46)
Research on the Integrated Science Curriculum	RISC	Personal gains in context of science	Attitudes about science – interdisciplinary		Likert-type	Online pre/post	Adaptable	Easy – done by author	
Science Motivation Questionnaire II	SMQII	Personal gains in context of science	Motivation to learn science	University of Georgia	Likert-type	In class or online	25 items	Easy	(29)
Survey of Undergraduate Research Experiences	SURE	Personal gains in context of science	Personal gains (UREs)	Multiple UG populations	Likert-type	Online pre/post	44 items	Easy – done by author	(45)
Undergraduate Student Self-Assessment Instrument	URSSA	Personal gains in context of science	Personal gains in context of science (UREs)	Multiple UG populations	Likert-type	Online post	Adaptable	Easy	(65)
Student Assessment of Learning Gains	SALG	Personal gains in context of science	Student perception of inquiry labs	Multiple UG populations	Likert-type	Online	5 items	Easy	(55)
Molecular Biology Data Analysis Test	MBDAT	Process of science	Data analysis	Multiple UG populations	Multiple choice	In class and online pre/post	20 items		(52)
Results Analysis Concept Inventory	RACI	Process of science	Data analysis	University of British Columbia	Multiple choice	In class pre/post	12 items	Easy	
Biological Experimental Design Concept Inventory	BEDCI	Process of science	Experimental design	University of British Columbia	Multiple choice	In class pre/post	14 items, 18 minutes	Easy	(23)
Expanded Experimental Design Ability Test	E-EDAT	Process of science	Experimental design	University of Washington	Open-ended	In class and online pre/post		Moderate – rubric	(15)
Experimental Design – First Year Undergraduate		Process of science	Experimental design	University of British Columbia	Multiple choice	In class pre/post	18 items	Easy	
Experimental Design Ability Test	EDAT	Process of science	Experimental design	Bowling Green State	Open-ended	In class pre/post	10-12 min	Moderate – rubric	(58)
Rubric for Experimental Design	RED	Process of science	Experimental design	Midwestern Research University	Writing sample	Pre-post online	NA	Moderate – rubric	(21)
Test of Scientific Literacy Skills	TOSLS	Process of science	Scientific literacy	Multiple populations	Multiple choice	In class pre/post	30 min	Easy	(30)
Classroom Test of Scientific Reasoning	CTSR	Process of science	Scientific reasoning	Multiple populations	Open-ended	Multiple choice	13 items	Moderate	(43)

TABLE I.
Continued.

Assessment Instrument Name	Acronym	Primary Aim	Secondary Aim	Validation Population	Answer Type	How Administered	Number of Items / Time to Administer	Scoring	Citation
The Rubric for Science Writing	Rubric	Process of science	Scientific reasoning / science communication	University of South Carolina	Writing-sample	Out of class	NA	Moderate – rubric	(62)
National Survey of Student Engagement	NSSE ^a	Student engagement	Student engagement	Multiple ages & populations	Likert-type	Online	70 items	Easy	(41)

^a Indicates that the instrument has a fee for use; ^b Instrument is to be used in conjunction with the POS. Blank cells indicate that the information was not specified.

NA = not applicable; UG = undergraduate; URE = undergraduate research experience.

steps meet minimum expectations for using best practices in evaluating CURE outcomes.

- How is this instrument aligned with the learning goals of my CURE? Is it specifically aimed at measuring this particular outcome?
- What populations has it been previously administered to? Does that student population reasonably match mine?
- What is the time needed to use, administer, and analyze the results of the instrument? Is this feasible within my course timeline and personal availability?

Possible follow-up question:

- Do I aim to use the assessment results outside of my own classroom and/or try to publish them? If no, then validity and reliability measures may be less critical for an interpretation of the results. If yes, what validity and reliability measures have been performed and reported on for this instrument, and should I consider collaborating with an education researcher?

Assessing student understanding of experimental design. To help instructors determine how to assess their CUREs, we have identified one of the most commonly expected learning outcomes from CUREs: *Students will learn how to design scientific experiments*. We conducted phone surveys with faculty members who we had previously interviewed regarding their experiences developing and teaching their own CURE (55). We asked them to identify whether they thought students gained particular outcomes as a result of participating in their CURE (See Appendix I for details). Of the 35 surveys conducted, 86% of faculty participants reported that they *perceived* that students learned to design scientific experiments as a result of the CURE. Using the steps outlined in this essay, we provide an example of how to begin to assess this learning outcome by considering the pros and cons of different assessment instruments (all cited in Table I). The instruments we discuss below have the explicit primary aim of evaluating student understanding of the “Process of Science” and the secondary aim of evaluating student understanding of “Experimental Design.” Further, the instruments were developed using undergraduate students at large, public research universities.

One of the first instruments to be developed to measure students’ ability to design biology experiments was the Experimental Design Ability Tool (EDAT) (56). The EDAT is a pre-post instrument, intended to be administered at the beginning and end of a course or module to evaluate gains in student ability. The EDAT consists of open-ended prompts asking students to design an experiment: the pretest prompt is focused on designing an investigation into the benefits of ginseng supplements, and the posttest prompt asks students to design an investigation into the

impact of iron supplements on women's memory. Student written responses to both prompts are evaluated using a rubric. This assessment was developed using a nonmajors biology class and has been since adapted for a majors class; the revised instrument is the Expanded-Experimental Design Ability Tool (E-EDAT) (14). The E-EDAT has the advantage that the revised rubric gives a more detailed report of student understanding, as it allows for intermediate evaluation of student ability to design experiments. However, the open-ended format of both these assessments means that grading student responses using the designated rubrics may be too time-consuming for many instructors. Additionally, the prompts of the EDATs are specific to human-focused medical scenarios, which may not reflect the type of experimental design that students are learning in their CURE.

Another pre-post assessment instrument, the Rubric for Experimental Design (RED), is a way to measure changes in student conceptions about experimental design (20). The RED is a rubric that can be used to evaluate student writing samples on experimental design, but is not associated with specific predetermined questions (20). Since many CUREs adopt a model where students write a final paper taking the form of a grant proposal or journal article, and the RED requires the instructor to have some sort of student writing sample already in place, the RED may be appropriate. Yet, similar to the EDAT/E-EDAT, the scoring of this instrument is time-consuming and the writing samples will need to be coded by more than one rater to achieve inter-rater reliability, which may be a limitation for some instructors. However, instructors using the RED have the advantage of a rubric that targets five common areas where students traditionally struggle regarding experimental design, thus potentially helping an instructor to disaggregate specific areas of student misconceptions and understanding of experimental design principles.

A pre-post, multiple-choice concept inventory, the Biological Experimental Design Concept Inventory (BEDCI), was developed to test student ability to design experiments (21). The BEDCI has the advantage that it is easy to score since it can be automated and the instructor can quickly identify student gains on the test, but a disadvantage is that the BEDCI consists of a fixed set of questions. The specific context of each question could impact how students perform on the assessment, and the context of these questions may not overlap with the context of the CURE. Additionally, the BEDCI is to be presented as a PowerPoint during class, so instructors need to allocate in-class time for administration.

These instruments may help an instructor to understand whether their students have achieved some level of experimental design ability, but the majority of these instruments are not specific to the context of any given CURE. Thus, there may be specific learning goals related to the experimental design context of the particular CURE that an instructor wants to probe. An additional and/or

alternative approach is to design a test of experimental design ability using the specific context of the CURE. While we often use the term "experimental design" to include any aspect of designing an experiment in science, aspects of experimental design in a molecular biology CURE are different than aspects of experimental design in a field ecology CURE. Further, even if students can design an experiment in one context, this does not mean that they can design an appropriate experiment in another context, nor should they necessarily be expected to do so, particularly if understanding nuances of both experimental systems was not a predetermined learning goal. Instructors may miss important gains in their students' abilities to design relevant experiments if they are using a generic experimental design assessment instrument (38). Perhaps students can design experiments in the specific context of their CURE (e.g., design an experiment to test the levels of protein in yeast cells in a molecular biology CURE versus design an experiment to identify abiotic factors influencing the presence of yeast in a flowering plant's nectar in an ecology CURE), but they are unable to effectively design experiments in the converse scientific context. Even skilled scientists can have difficulty in designing an experiment in an area that is not in their specific domain of biological expertise. It may be important to test students using their specific CURE context in order to maximize the chance of seeing an outcome effect that can be credibly attributed to the CURE. It is unlikely that a previously developed assessment instrument will be directly aligned with expected outcomes from one's CURE, so we encourage instructors to work with education researchers to develop situated assessments that are appropriate for each specific CURE context (e.g., 38).

CONCLUSION

As more CUREs are developed and implemented in biology lab courses across the country, instructors are becoming increasingly interested in assessing the impact of their CUREs. Although there is complexity in assessment, the aim of this paper is not to overwhelm instructors, but instead to offer a basic assessment strategy: identify anticipated CURE learning outcomes, select an assessment instrument that is aligned with the learning outcomes, and cautiously interpret the results of the assessment instrument. We also present a table of previously developed assessment instruments that could be of use to CURE instructors depending on their learning goals, student populations, and course context. While this is only the tip of the iceberg as far as how instructors can assess their CUREs, and we anticipate that many more assessment instruments will be developed in the coming years, we hope that this table can provide instructors with a starting point for considering how to assess their CUREs. We encourage instructors to be thoughtful and critical in their assessment of their CUREs as we continue to learn more about the impact of these curricula on students.

SUPPLEMENTAL MATERIALS

Appendix I: Faculty perceptions of student gains from participation in CUREs

ACKNOWLEDGMENTS

The authors would like to thank J. Barbera, K. McDonald, and K. Stedman for their thoughtful comments on an early version of the manuscript. We also thank the faculty members who participated in the study referenced in the Appendix. The authors declare that there are no conflicts of interest.

REFERENCES

1. **American Association for the Advancement of Science.** 2011. Vision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009. Washington, DC.
2. **American Educational Research Association, American Psychological Association, National Council on Measurement in Education.** 1999. Standards for educational and psychological testing. AERA Publication Series, Washington, DC.
3. **Auchincloss, L. C., et al.** 2014. Assessment of course-based undergraduate research experiences: a meeting report. *CBE Life Sci. Educ.* **13**:29–40.
4. **Bangera, G., and S. E. Brownell.** 2014. Course-based undergraduate research experiences can make scientific research more inclusive. *CBE Life Sci. Educ.* **13**:602–606.
5. **Barbera, J., and J. VandenPlas.** 2011. All assessment materials are not created equal: the myths about instrument development, validity, and reliability, p 177–193. *In*: Bunce, D. M. (ed.), *Investigating Classroom Myths through Research on Teaching and Learning*. American Chemical Society, Washington, DC.
6. **Bascom-Slack, C. A., A. E. Arnold, and S. A. Strobel.** 2012. Student-directed discovery of the plant microbiome and its products. *Science* **338**:485–486.
7. **Biggs, J., D. Kember, and D. Y. Leung.** 2001. The revised two-factor study process questionnaire: R-SPQ-2F. *Brit. J. Educ. Psychol.* **71**:133–149.
8. **Bollen, K. A., and R. H. Hoyle.** 1990. Perceived cohesion: a conceptual and empirical examination. *Soc. Forces* **69**(2):479–504.
9. **Bowman, N. A.** 2011. Validity of college self-reported gains at diverse institutions. *Educ. Res.* **40**:22–24.
10. **Brownell, S. E., et al.** 2015. A high-enrollment course-based undergraduate research experience improves student conceptions of scientific thinking and ability to interpret data. *CBE Life Sci. Educ.* **14**:ar21.
11. **Brownell, S. E., and M. J. Kloser.** 2015. Toward a conceptual framework for measuring the effectiveness of course-based undergraduate research experiences in undergraduate biology. *Studies Higher Educ.* **40**:525–544.
12. **Brownell, S. E., M. J. Kloser, T. Fukami, and R. Shavelson.** 2012. Undergraduate biology lab courses: comparing the impact of traditionally based “cookbook” and authentic research-based courses on student lab experiences. *J. Coll. Sci. Teach.* **41**:18–27.
13. **Brownell, S. E., M. J. Kloser, T. Fukami, and R. J. Shavelson.** 2013. Context matters: volunteer bias, small sample size, and the value of comparison groups in the assessment of research-based undergraduate introductory biology lab courses. *J. Microbiol. Biol. Educ.* **14**(2):176–182.
14. **Brownell, S. E., et al.** 2013. How students think about experimental design: novel conceptions revealed by in-class activities. *BioScience*:bit016.
15. **Chan, D.** 2009. So why ask me? Are self-report data really that bad?, p 309–336. *In*: Lance, C. E., and R. J. Vandenberg (eds.), *Statistical and methodological myths and urban legends: doctrine, verity and fable in the organizational and social sciences*. Routledge, New York, NY.
16. **Corwin, L. A., M. J. Graham, and E. L. Dolan.** 2015. Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. *CBE Life Sci. Educ.* **14**:es1.
17. **Corwin, L. A., C. Runyon, A. Robinson, and E. L. Dolan.** 2015. The laboratory course assessment survey: a tool to measure three dimensions of research-course design. *CBE Life Sci. Educ.* **14**:ar37.
18. **Creamer, E. G., M. B. Magolda, and J. Yue.** 2010. Preliminary evidence of the reliability and validity of a quantitative measure of self-authorship. *J. Coll. Student Devt.* **51**:550–562.
19. **Crowe, A., C. Dirks, and M. P. Wenderoth.** 2008. Biology in bloom: implementing Bloom’s taxonomy to enhance student learning in biology. *CBE Life Sci. Educ.* **7**:368–381.
20. **Dasgupta, A. P., T. R. Anderson, and N. Pelaez.** 2014. Development and validation of a rubric for diagnosing students’ experimental design knowledge and difficulties. *CBE Life Sci. Educ.* **13**:265–284.
21. **Deane, T., K. Nomme, E. Jeffery, C. Pollock, and G. Birol.** 2014. Development of the biological experimental design concept inventory (BEDCI). *CBE Life Sci. Educ.* **13**:540–551.
22. **Dolan, E. L., and J. P. Collins.** 2015. We must teach more effectively: here are four ways to get started. *Mol. Biol. Cell* **26**:2151–2155.
23. **Duckworth, A. L., and P. D. Quinn.** 2009. Development and validation of the short grit scale (GRIT-S). *J. Pers. Assess.* **91**:166–174.
24. **Dunlap, R., K. Liere, A. Mertig, and R. E. Jones.** 2000. Measuring endorsement of the new ecological paradigm: a revised NEP scale. *J. Soc. Iss.* **56**:425–442.
25. **Ebel, R., and D. Frisbie.** 1979. *Essentials of educational measurement*. Prentice-Hall, Englewood Cliffs, NJ.
26. **Facione, P. A.** 1991. Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment. [Online.]

- <http://www.insightassessment.com/Products/Products-Summary/Critical-Thinking-Skills-Tests/California-Critical-Thinking-Skills-Test-CCTST>.
27. Glynn, S. M., P. Brickman, N. Armstrong, and G. Taasobshirazi. 2011. Science motivation questionnaire II: validation with science majors and nonscience majors. *J. Res. Sci. Teach.* **48**:1159–1176.
 28. Gormally, C., P. Brickman, and M. Lutz. 2012. Developing a test of scientific literacy skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE Life Sci. Educ.* **11**:364–377.
 29. Gurel, D. K., A. Eryilmaz, and L. C. McDermott. 2015. A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia J. Math. Sci. Technol. Educ.* **11**:989–1008.
 30. Halloun, I., and D. Hestenes. 1998. Interpreting VASS dimensions and profiles for physics students. *Sci. Educ.* **7**:553–577.
 31. Hanauer, D. I., and E. L. Dolan. 2014. The project ownership survey: measuring differences in scientific inquiry experiences. *CBE Life Sci. Educ.* **13**:149–158.
 32. Hanauer, D. I., and G. Hatfull. 2015. Measuring networking as an outcome variable in undergraduate research experiences. *CBE Life Sci. Educ.* **14**:ar38.
 33. Handelsman, J., et al. 2004. Scientific teaching. *Science* **304**:521–522.
 34. Hark, A. T., C. P. Bailey, S. Parrish, W. Leung, C. D. Shaffer, and S. C. Elgin. 2011. Undergraduate research in the genomics education partnership: a comparative genomics project exploring genome organization and chromatin structure in *Drosophila*. *FASEB J.* **25**:576.573.
 35. Harrison, M., D. Dunbar, L. Ratmanský, K. Boyd, and D. Lopatto. 2011. Classroom-based science research at the introductory level: changes in career choices and attitude. *CBE Life Sci. Educ.* **10**:279–286.
 36. Jordan, T. C., et al. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *MBio* **5**:e01051–01013.
 37. Kloser, M. J., S. E. Brownell, N. R. Chiariello, and T. Fukami. 2011. Integrating teaching and research in undergraduate biology laboratory education. *PLoS Biol.* **9**:e1001174.
 38. Kloser, M. J., S. E. Brownell, R. J. Shavelson, and T. Fukami. 2013. Effects of a research-based ecology lab course: a study of nonvolunteer achievement, self-confidence, and perception of lab course purpose. *J. Coll. Sci. Teach.* **42**:90–99.
 39. Kuh, G. D. 2009. The national survey of student engagement: conceptual and empirical foundations. *New Direct. Inst. Res.* **2009**:5–20.
 40. Kuncel, N. R., M. Credé, and L. L. Thomas. 2005. The validity of self-reported grade point averages, class ranks, and test scores: a meta-analysis and review of the literature. *Rev. Educ. Res.* **75**:63–82.
 41. Lawson, A. E., B. Clark, E. Cramer-Meldrum, K. A. Falconer, J. M. Sequist, and Y. J. Kwon. 2000. Development of scientific reasoning in college biology: do two levels of general hypothesis-testing skills exist? *J. Res. Sci. Teach.* **37**:81–101.
 42. Lederman, N. G., F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz. 2002. Views of nature of science questionnaire: toward valid and meaningful assessment of learners' conceptions of nature of science. *J. Res. Sci. Teach.* **39**:497–521.
 43. Lopatto, D. 2004. Survey of undergraduate research experiences (SURE): first findings. *Cell Biol. Educ.* **3**:270–277.
 44. Lopatto, D. 2008. Classroom Undergraduate Research Experiences Survey (CURE). [Online.] <https://www.grinnell.edu/academics/areas/psychology/assessments/cure-survey>.
 45. Lopatto D., et al. 2014. A central support system can facilitate implementation and sustainability of a classroom-based undergraduate research experience (CURE) in Genomics. *CBE Life Sci. Educ.* **13**:711–723.
 46. Makarevitch, I., C. Frechette, and N. Wiatros. 2015. Authentic research experience and “big data” analysis in the classroom: maize response to abiotic stress. *CBE Life Sci. Educ.* **14**.
 47. Messick, S. 1989. Meaning and values in test validation: the science and ethics of assessment. *Educ. Res.* **18**:5–11.
 48. Milfont, T. L., and J. Duckitt. 2010. The environmental attitudes inventory: a valid and reliable measure to assess the structure of environmental attitudes. *J. Environ. Psychol.* **30**:80–94.
 49. President's Council of Advisors on Science and Technology. 2012. Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. PCAST, Washington, DC.
 50. Rybarczyk, B. J., K. L. Walton, and W. H. Grillo. 2014. The development and implementation of an instrument to assess students' data analysis skills in molecular biology. *J. Microbiol. Biol. Educ.* **15**(2):259–267.
 51. Sanders, E. R., and A. M. Hirsch. 2014. Immersing undergraduate students into research on the metagenomics of the plant rhizosphere: a pedagogical strategy to engage civic-mindedness and retain undergraduates in STEM. *Front. Plant Sci.* **5**:157.
 52. Semsar, K., J. K. Knight, G. Birol, and M. K. Smith. 2011. The Colorado learning attitudes about science survey (CLASS) for use in biology. *CBE Life Sci. Educ.* **10**:268–278.
 53. Seymour, E., D. Wiese, A. Hunter, and S. M. Daffinrud. 2000. Creating a better mousetrap: on-line student assessment of their learning gains. Paper presented at the National Meeting of the American Chemical Society, San Francisco, CA.
 54. Shaffer, C. D., et al. 2014. A course-based research experience: how benefits change with increased investment in instructional time. *CBE Life Sci. Educ.* **13**:111–130.
 55. Shortlidge, E. E., G. Banger, and S. E. Brownell. 2016. Faculty perspectives on developing and teaching course-based undergraduate research experiences. *BioScience* **66**:54–62.

56. **Sirum, K., and J. Humburg.** 2011. The experimental design ability test (EDAT). *Bioscene J. Coll. Biol. Teach.* **37**:8–16.
57. **Stein, B., A. Haynes, M. Redding, T. Ennis, and M. Cecil.** 2007. Assessing critical thinking in STEM and beyond, p 79–82. *In: Innovations in e-learning, instruction technology, assessment, and engineering education.* Springer, Netherlands.
58. **Stiggins, R. J.** 2004. Classroom assessment for student learning: doing it right—using it well. Assessment Training Institute Inc., Portland, OR.
59. **Theobald, R., and S. Freeman.** 2014. Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE Life Sci. Educ.* **13**:41–48.
60. **Timmerman, B. E C., D. C. Strickland, R.L. Johnson, and J. R. Payne.** 2011. Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assess. Eval. Higher Educ.* **36**:509–547.
61. **Torrance, E. P.** 1998. Torrance tests of creative thinking: norms-technical manual: figural (streamlined) forms A & B. Scholastic Testing Service Inc., Bensenville, IL.
62. **Ward, J. R., H. D. Clarke, and J. L. Horton.** 2014. Effects of a research-infused botanical curriculum on undergraduates' content knowledge, STEM competencies, and attitudes toward plant sciences. *CBE Life Sci. Educ.* **13**:387–396.
63. **Weston, T. J., and S. L. Laursen.** 2015. The undergraduate research student self-assessment (URSSA): validation for use in program evaluation. *CBE Life Sci. Educ.* **14**(3). pii:ar33.
64. **Wiggins, G. P., J. McTighe, L. J. Kiernan, and F. Frost.** 1998. Understanding by design. Association for Supervision and Curriculum Development, Alexandria, VA.